

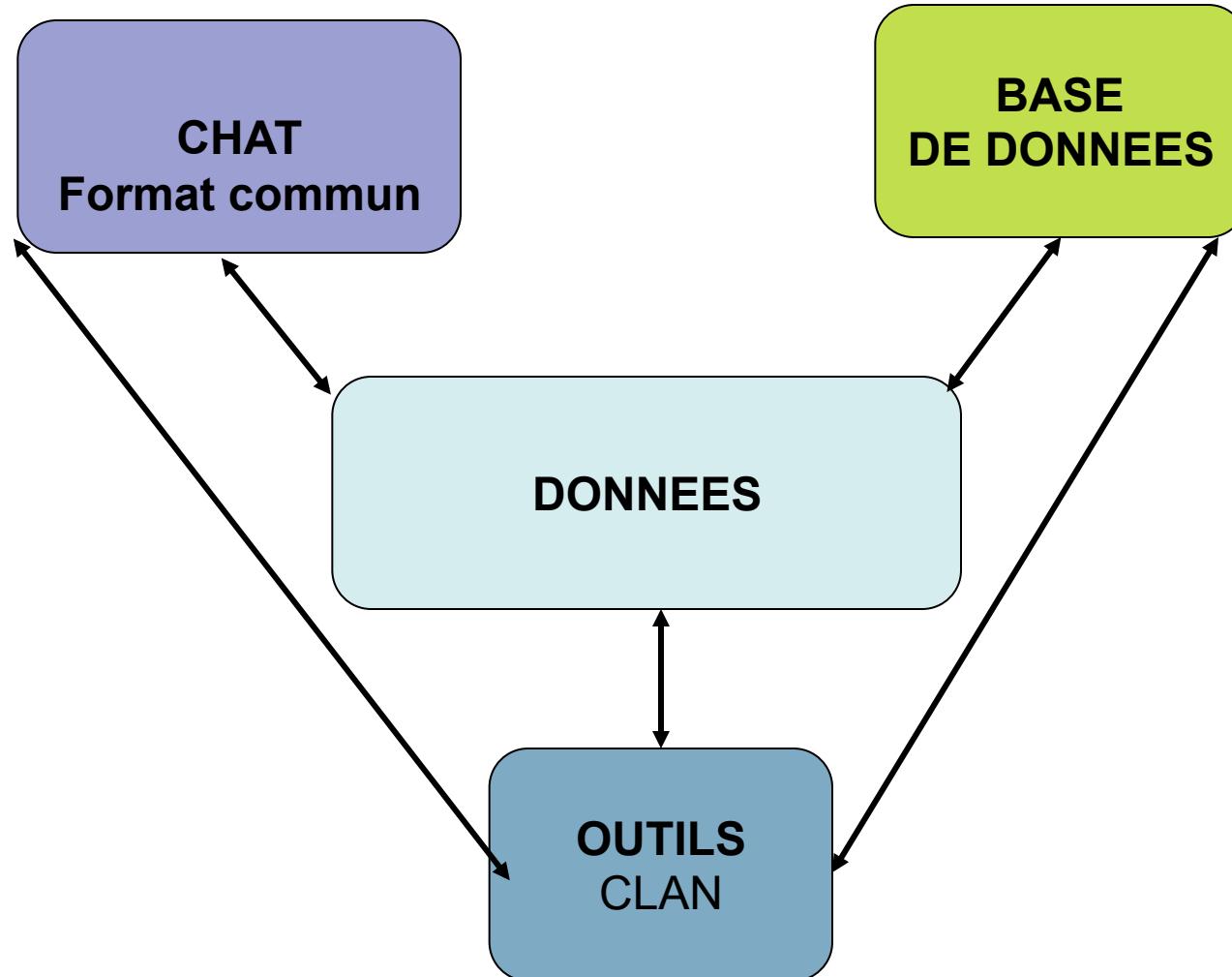
CHILDES

(Child Language Data Exchange System)

Ensemble d' éléments permettant à la communauté scientifique d'échanger des corpus de langage d'enfants

Créé en 1984 par [Brian MacWhinney](#) et [Catherine Snow](#) (Université Carnegie Mellon de Pittsburgh, USA)

CHILDES



Type de données

- actuellement
 - des transcriptions simples (textes seulement, avec ou sans phonétique)
 - utilisation de la phonétique (y compris au format API), lien avec le son, analyse des fichiers sons
 - liens avec la vidéo
- futur (améliorations en cours)
 - nouvelle base de données : TALKBANK,
 - nouveaux outils : meilleure analyse des sons (phonèmes), des images (gestes)

CHAT

- au départ: texte brut qui peut être édité et manipulé avec n'importe quel logiciel
- plus tout à fait vrai depuis qu'il est possible d'insérer du son ou de la vidéo, ainsi que d'utiliser des symboles API pour représenter la phonétique

Principes de CHAT

Il y a trois types d'information stockés dans une description

- 1) des informations à caractère général qui se rapportent à tout l'enregistrement
- 2) des transcriptions réalisées énoncé par énoncé (ou par tours de parole)
- 3) des indications complémentaires se rapportant à un énoncé ou à un tour de parole précis

Premier niveau de format

- le type d'information est indiqué par le premier caractère d'une ligne

— @ informations de type général

— * énoncé (transcription « principale » le plus souvent en caractères graphémiques) – **lignes principales**

— % indications complémentaires correspondant à la ligne * située juste au-dessus dans la transcription – **lignes dépendantes**

— **tabulation** → continuation de la ligne du dessus lorsqu'elle ne peut tenir sur une seule ligne

Premier niveau de format

EXAMPLE

@Participants: CHI Ross
Enfant,

FAT Brian Père

*CHI: tu veux que je le ferme
ton sac ?

%pho: ty və kə ʒə lə feym t̪
sak

Deuxième niveau de format

- chacun des caractères @, * et % est suivi d' un mot indiquant la signification de la ligne, suivi d' un « : » et d' une tabulation
- pour * et %, le « mot » qui suit doit être un code d' exactement 3 caractères

Respect des conventions

- peu de marge d' erreur dans le codage (voir instrument CHECK)
- surtout faire attention aux espaces et tabulations :
 - pas d' espace avant @, *, %
 - pas d' espace après @, *, % et avant le mot qui suit (*CHI:)
 - pas d' espace après le mot qui suit et avant le : (*CHI:)
 - ne pas oublier le caractère : et la tabulation qui suit
 - ne pas oublier de mettre une tabulation pour continuer une ligne

Organisation d' une transcription

```
@Begin
@Languages: fr
@Participants: CHI Charles Child, MOT Mère Mother
@ID: fr|exemple|CHI||3;06.00|male|group|ses|Child|mat|
@ID: fr|exemple|MOT||female|group|ses|Mother|high|
*CHI: c'+est maman.
*MOT: c'+est moi, ça?
*CHI: oui.
*CHI: xx jouer.
*MOT: tu joues?
%com: répétition de la mère
*CHI: moi je jouer.
@End
```

Codes obligatoires

- @Begin
- @Languages: suivi d' un code langage
- @Participants: liste des participants
- @ID: identification d' un des participants
- ...
- @End

Codes obligatoires

- **@Participants:**
 - élément de la liste en trois parties
 - code de trois lettres majuscules
 - optionnel: nom de l' enfant, indication utile
 - role: indication du rôle du locuteur dans la situation → limité à une liste (très large) comprenant Target_Child, Child, Mother, ..., Investigator, ..., Camera_Operator, ..., Unidentified, ...

Quelques codes optionnels en début de corpus

@Birth of XXX: 01-01-1999

@Coder: nom de la personne qui transcrit
partout dans le corpus

@Comment:

@Date:

@Location:

@Situation:

Codages des transcriptions

- les transcriptions principales doivent se terminer par une ponctuation . ? ! ou un symbole spécial indiquant un énoncé inachevé +... ou interrompu +/.
- on peut aussi utiliser des marqueurs de fin caractérisant le contour intonatif
 - ? montant -! exclamatif
 - . descendant - ' . montant-
 - , . descendant-montant

Codages des mots

- Un mot se code normalement en signe alphabétique
→ **maman**
- Les apostrophes doivent être suivies d'un espace
→ **d' abord**
- Les ‘-’ des mots composés sont remplacés par des ‘+’ → **pommes+de+terre**
- Les ‘-’ des critiques post-verbaux sont supprimés
→ **donne le**

Codages des mots

- les mots incomplets peuvent être complétés → **(pe)tit (en)core**
- les pauses peuvent être marquées → #
- les reprises peuvent codées
 - <le pe> [/] le petit bébé
 - <le pe> [//] les petits békés
 - <le pe> [///] un garçon
- le mot précédent est incertain → [?]
- le mot précédent est erroné → [*]

Codages des mots

- les mots peuvent complétés par des indicateurs → ding@o
- un codage phonétique optionnel est codé comme → ſeſe@u
- les énoncés incompréhensibles sont codés comme xxx → *CHI: xxx.
- les mots incompréhensibles sont codés comme xx → *CHI: xx tombe.

Autres codages

- superposition d' énoncés
- continuation de l' énoncé précédent
- variété de pauses → #d #2.3
- acronymes → w_c
- lettres isolées → a@l
- omission de mots
- commentaires dans une ligne → [= xy]